

Tijdreeksen

Onderzoekstechnieken

Jens Buysse Wim De Bruyn Bert Van Vreckem
AJ 2018-2019

What's on the menu today?

- Tijdreeksen en voorspellingen

- Tijdreeksmodellen

 - Wiskundig model

 - Algemeen

 - Schatten van de parameters

- Voortschrijdend gemiddelde

 - Eenvoudig voortschrijdend gemiddelde

 - Gewogen voortschrijdend gemiddelde

 - Dubbele exponentiële afvlakking

 - Voorspellen

 - Driedubbele exponentiële afvlakking

Tijdreeksen en voorspellingen

Tijdreeksen en voorspellingen

Een tijdreeks is een opeenvolging van observaties van een variabele in functie van de tijd

Een tijdreeks is een stochastisch proces. Voorbeelden:

- Maandelijks vraag naar melk
- Jaarlijkse instroom van “generatiestudenten” aan de hogeschool
- Prijs van een aandeel of obligatie op de beurs (van uur tot uur, per dag, ...)
- Aantal HTTP requests per seconde op een website
- Evolutie schijfgebruik op een backup-server

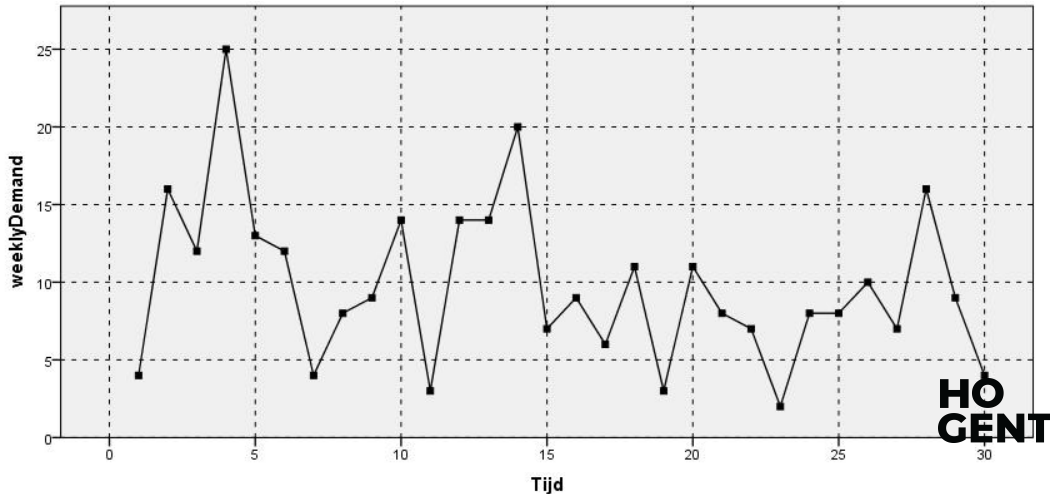
Tijdreeksen en voorspellingen

Tijdreeksen zijn een belangrijk onderdeel van onderzoek omdat ze vaak de **basis** vormen voor beslissingsmodellen en voorspellingen.

- algemene ontwikkeling van toekomstplannen (investeringen, capaciteit ...)
- plannen van budgettering om tekortkomingen te vermijden (operationeel budget, marketing budget ...)
- ondersteuning van financiële objectieven
- onzekerheid vermijden

Tijdreeksen en voorspellingen

Tijdreeksen zijn een **statistisch** probleem: observaties variëren in functie van de tijd



Tijdreeksmodellen

Wiskundig model tijdreeks

Het eenvoudigste model:

- Constante b
- Variaties rond b bepaald door willekeurige variabele ε_t

$$X_t = b + \varepsilon_t \quad (1)$$

- X_t : stochastische variabele voor de tijdreeks, op tijdstip t
- x_t : *observatie* op tijdstip t (dus gekend!)
- ε_t noemt men de *storing* (Eng: *noise*). We veronderstellen $\varepsilon_t \sim \text{Nor}(\mu = 0; \sigma)$

Wiskundig model tijdreeks

We kunnen er ook van uit gaan dat er een lineair verband is:

$$X_t = b_0 + b_1 \times t + \varepsilon_t \quad (2)$$

Vergelijkingen 1 en 2 zijn speciale gevallen van het *polynomiale* geval:

$$X_t = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n + \varepsilon_t \quad (3)$$

Algemene uitdrukking tijdreeks

$$X_t = f(b_0, b_1, b_2, \dots, b_n, t) + \varepsilon_t \quad (4)$$

We gaan verder uit van deze veronderstellingen:

- We beschouwen twee componenten van variabiliteit:
 - het gemiddelde van de voorspellingen verandert in de tijd
 - de variaties ten opzichte van dit gemiddelde variëren willekeurig
- De variatie van de residuen van het model ($X_t - x_t$) is constant in de tijd (*homoscedastisch*)

Schatten van de parameters

Voorspellingen maken aan de hand van tijdreeksmodel:

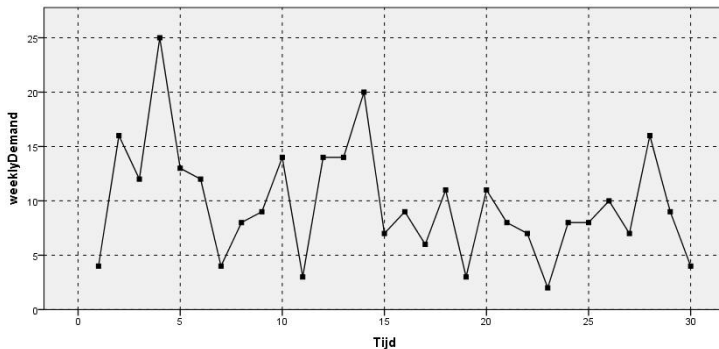
1. selecteer het meest passende model
2. schatting voor parameters $b_i (i : 1, \dots, n)$ a.h.v. observaties

Deze schattingen \hat{b}_i zijn dan zodanig dat ze de geobserveerde waarden zo goed mogelijk benaderen.

Voorbeeld

| | | | | | | | | | |
|---|----|----|----|----|----|---|----|---|----|
| 4 | 16 | 12 | 25 | 13 | 12 | 4 | 8 | 9 | 14 |
| 3 | 14 | 14 | 20 | 7 | 9 | 6 | 11 | 3 | 11 |
| 8 | 7 | 2 | 8 | 8 | 10 | 7 | 16 | 9 | 4 |

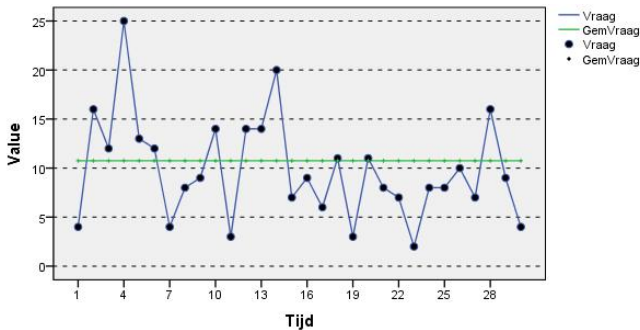
Tabel: Tijdreeks voor de wekelijkse vraag voor een product



Voorbeeld: Parameterschatting

- We kiezen het constante model uit vergelijking 1
- Als schatter voor b kiezen we het gemiddelde van de eerste 20 observaties:

$$\hat{b} = \frac{1}{20} \sum_{t=1}^{20} x_t = 10.75$$



Voorbeeld: parameterschatting

Je kiest zelf welke observaties je gebruikt voor het bepalen van \hat{b} , bv.:

- $\hat{b} = \frac{1}{10} \sum_{10}^{20} x_t = 10.18$
- $\hat{b} = \frac{1}{5} \sum_{15}^{20} x_t = 7.83$

Voortschrijdend gemiddelde

Voortschrijdend gemiddelde

Het eenvoudig voortschrijdend gemiddelde is een reeks *gemiddelden* van de laatste m observaties

- Eng.: *Simple Moving Average* (SMA)
- Verbergen korte-termijn fluctuaties en tonen lange-termijn trends
- m is het tijdbereik (time range), en is de parameter van deze methode

$$SMA(t) = \sum_{i=k}^t \frac{x_i}{m} \quad (5)$$

met $k = t - m + 1$.

Voorbeeld: "Golden cross"

Moving Averages worden gebruikt in *technische analyse* van aandelenkoersen om trends te ontdekken:



OZT: Tijdreeksen

- └ Voortschrijdend gemiddelde
- └ Eenvoudig voortschrijdend gemiddelde
- └ Voorbeeld: “Golden cross”

Voorbeeld: “Golden cross”

Moving Averages worden gebruikt in technische analyse van aandelenkoersen om trends te ontdekken:



Deze grafiek is de koers van de S&P500 index (vergelijk met onze Bel-20) voor december 2011 tot eind 2012. In de grafiek worden de SMA's van de prijs bij sluitingstijd van de 50 en 200 laatste beursdagen gegeven (notatie: MA(50) en MA(200)). Wanneer de markt in een dalende trend zit (“**bear market**”), ligt de MA(200) boven MA(50) (zie linkse deel van de tekening).

In februari 2012 ging MA(50) boven de MA(200), een verschijnsel dat “**the golden cross**” genoemd wordt. De stijging van de index was al enkele maanden eerder ingezet, maar de MA's reageren uiteraard trager.

Een *golden cross* is een indicator dat de markt (of een specifiek aandeel) in een lange-termijn stijgende trend zit (“**bull market**”). In dit geval duurt deze trend nog voort tot vandaag (voorjaar 2015).

De MA(200)-lijn wordt dan beschouwd als “steunpunt”, d.w.z. een ondergrens voor het koerscijfer. Aan deze prijs is het aandeel heel interessant, stijgt de vraag en wordt de koers terug omhoog getrokken.

| | | | | | | | | | |
|-------|------|------|------|------|------|------|------|------|------|
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| x_t | 3 | 14 | 14 | 20 | 7 | 9 | 6 | 11 | 3 |
| X_t | 11.7 | 11.6 | 11.4 | 11.6 | 11.1 | 10.5 | 10.2 | 10.4 | 10.7 |
| e | -8.7 | 2.4 | 2.6 | 8.4 | -4.1 | -1.5 | -4.2 | 0.6 | -7.7 |

Tabel: Voorspellingsfout voor een moving average $m = 10$

Een methode om de voorspelling te meten is het gemiddelde van de deviaties (*MAD*).

$$MAD = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (6)$$

Je kan ook de variantie ervan bepalen:

$$s_e^2 = \frac{1}{m} \sum_{i=1}^n (e_i - \bar{e})^2$$

Gewogen voortschrijdend gemiddelde

- Bij *SMA* zijn de gewichten van de observaties gelijk
- Bij gewogen voortschrijdend gemiddelde (weighted moving average, *WMA*) wegen recentere observaties meer door
- Een specifieke vorm hiervan is exponentiële afvlakking (exponential smoothing) of het *exponentieel voortschrijdend gemiddelde (EMA)*:

$$EMA(t) = \alpha x_{t-1} + (1 - \alpha)EMA(t - 1) \quad (8)$$

met α de smoothing constante ($0 < \alpha < 1$), en $t \geq 3$

Exponentiële afvlakking

Formule 8 geldt enkel vanaf $t = 3$. Het bepalen van $EMA(2)$ is een belangrijke parameter.

Er zijn verschillende keuzes:

- $EMA(2) = x_1$
- $EMA(2) = \frac{1}{m} \sum_{i=1}^m x_i$ (dus gemiddelde van de eerste m observaties)
- $EMA(2)$ gelijk stellen aan een bepaald objectief
- ...

Waarom “*exponentieel*”?

$$\begin{aligned}EMA(t) &= \alpha x_{t-1} + (1 - \alpha)EMA(t - 1) \\&= \alpha x_{t-1} + (1 - \alpha) [\alpha x_{t-2} + (1 - \alpha)EMA(t - 2)] \\&= \alpha x_{t-1} + \alpha(1 - \alpha)x_{t-2} + (1 - \alpha)^2EMA(t - 2) \\&\text{of algemeen gesteld:}\end{aligned}$$

$$= \alpha \sum_{i=1}^{t-2} (1 - \alpha)^{i-1} x_{t-i} + (1 - \alpha)^{t-2} EMA(2), t \geq 2$$

M.a.w. oudere observaties krijgen een exponentieel kleiner gewicht.

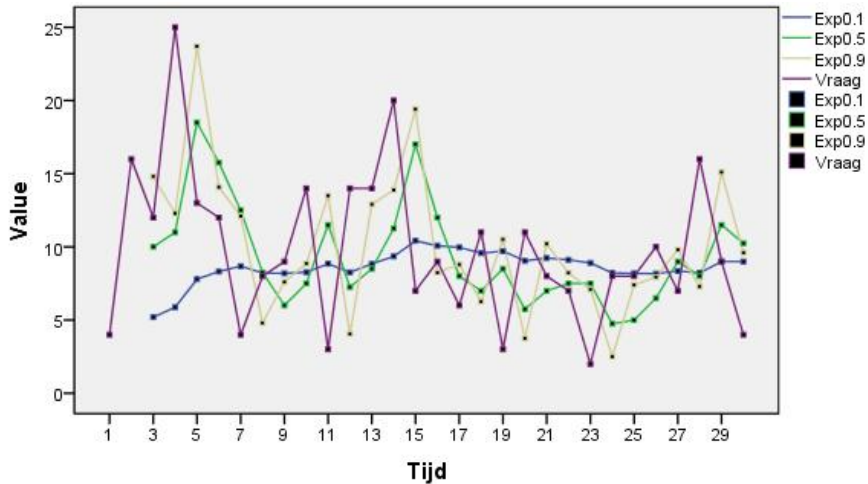
Exponentiële afvlakking

| α | $(1 - \alpha)$ | $(1 - \alpha)^2$ | $(1 - \alpha)^3$ | $(1 - \alpha)^4$ |
|----------|----------------|------------------|------------------|------------------|
| 0.9 | 0.1 | 0.01 | 0.001 | 0.0001 |
| 0.5 | 0.5 | 0.25 | 0.125 | 0.062 |
| 0.1 | 0.9 | 0.81 | 0.729 | 0.6561 |

Tabel: Waarden voor α en $(1 - \alpha)^n$

De snelheid waarmee de oude observaties "vergeten" worden hang af van α . Met een α dicht bij 1 vergeet je snel, terwijl een α dicht bij nul ervoor zorgt dat dat vergeten minder snel gaat

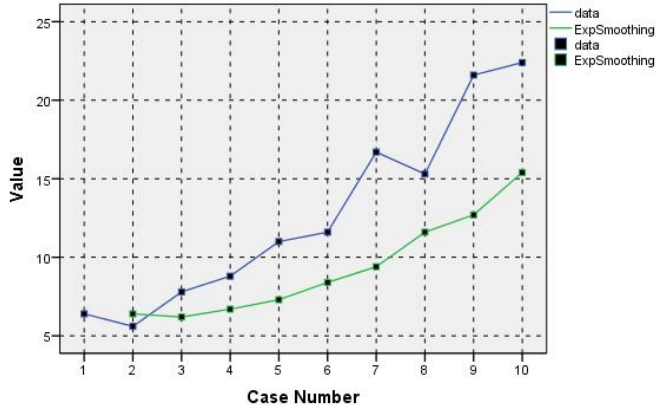
Voorbeeld



Figuur: Enkelvoudige exponentiële afvlakking met $\alpha = 0.1, 0.5, 0.9$

Dubbele exponentiële afvlakking

Enkelvoudige afvlakking werkt niet goed als er een trend in de data zit



Dubbele exponentiële afvlakking

We voeren een extra term in om de trend te modelleren. We noteren s_t voor de afgevlakte waarde, en b_t voor de schatting van de trend op tijdstip $t > 1$:

$$X_t = \alpha X_t + (1 - \alpha)(X_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1$$

$$b_t = \beta(X_t - X_{t-1}) + (1 - \beta)b_{t-1} \quad 0 \leq \beta \leq 1$$

met $0 < \alpha < 1$ en $0 < \beta < 1$

- De b_{t-1} in de eerste vergelijking zorgt voor het volgen van de trend
- $X_t - X_{t-1}$ is positief of negatief en zorgt ervoor dat de trend gegenereerd wordt

Dubbele exponentiële afvlakking

Er bestaan opnieuw verschillende alternatieven om de initiële waarden te kiezen:

$$X_1 = x_1$$

$$b_1 = x_2 - x_1$$

$$b_1 = \frac{1}{3} [(x_2 - x_1) + (x_1 - x_2) + (x_4 - x_3)]$$

$$b_1 = \frac{x_n - x_1}{n - 1}$$

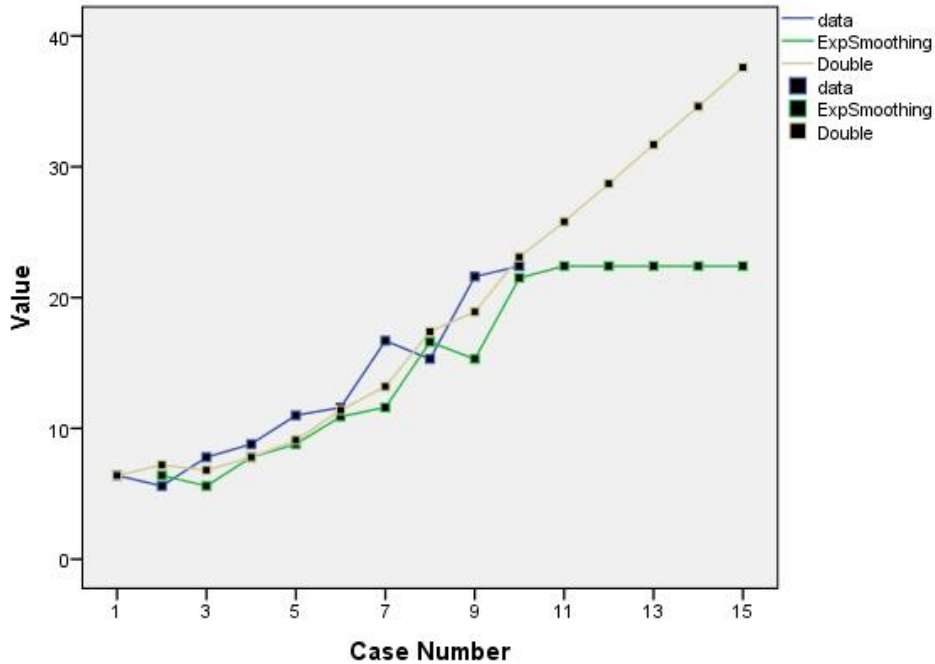
Voorspellen

Om een voorspelling $F(t + 1)$ te maken voor tijdstip $t + 1$ gebruiken we:

$$F(t + 1) = s_t + b_t$$

of algemeen voor tijdstip $t + m$:

$$F(t + m) = s_t + mb_t$$



**HO
GENT**

Driedubbele exponentiële afvlakking

of methode van Holt-Winters. Deze houdt rekening met seizoenaliteit in de data. We noteren:

- L : lengte van de seizoenale cyclus (aantal tijdseenheden)
- c_t : term die de seizoenale variaties modelleert
- γ : smoothing factor voor de seizoenale variatie

$$X_t = \alpha \frac{X_t}{c_{t-L}} + (1 - \alpha)(X_{t-1} + b_{t-1}) \quad \text{Smoothing}$$

$$b_t = \beta(X_t - X_{t-1}) + (1 - \beta)b_{t-1} \quad \text{Trend smoothing}$$

$$c_t = \gamma \frac{X_t}{X_t} + (1 - \gamma)c_{t-L} \quad \text{Seasonal smoothing}$$

Driedubbele exponentiële afvlakking

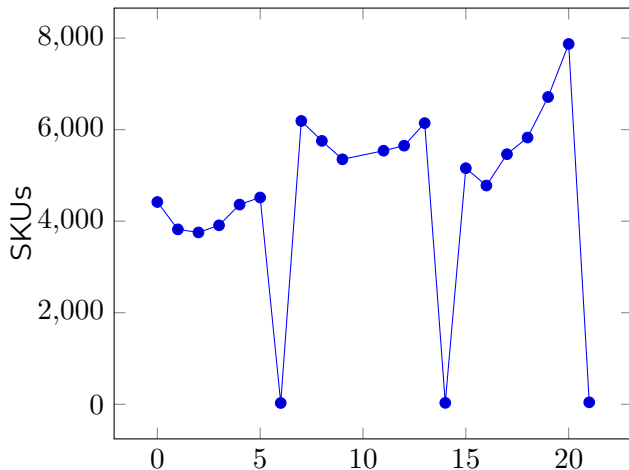
Voorspelling op tijdstip $t + m$:

$$F_{t+m} = (X_t + mb_t)c_{t-L+m} \text{Voorspelling}$$

Voor een implementatie in Java, zie
<https://github.com/bertvv/wintersmethod>

Voorbeeld: voorspel verkoopscijfers

Geobserveerde Verkoopscijfers



**HO
GENT**

Voorbeeld

Keuze startwaarden:

- smoothing factors: $\alpha = 0.8, \beta = 0.8, \gamma = 0.3$
- $s_0 = 5849, b_0 = 123.3$
- $L = 7$ (dus wekelijks terugkerend), dan hebben we ook 7 waarden nodig om c_t te initialiseren (zie tabel)

| | | | |
|--------------|--------------|--------------|--------------|
| ma (c_0) | di (c_1) | wo (c_2) | do (c_3) |
| 1.245693 | 1.115265 | 1.088853 | 1.135378 |
| vr (c_4) | za (c_5) | zo (c_6) | |
| 1.178552 | 1.229739 | 0.006520 | |

Tabel: Startwaarden voor c_t

Voorbeeld: voorspelling

